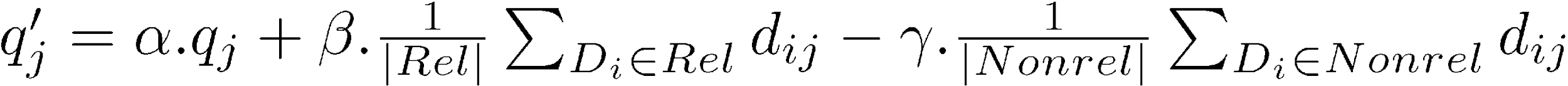
**Algorithm 1**

**Information filtering Rocchio model**

In this model we use the following formula;



qj is the initial query vector and qj’ is the updated query vector. I use the constant values;

alpha = 2

beta = 4

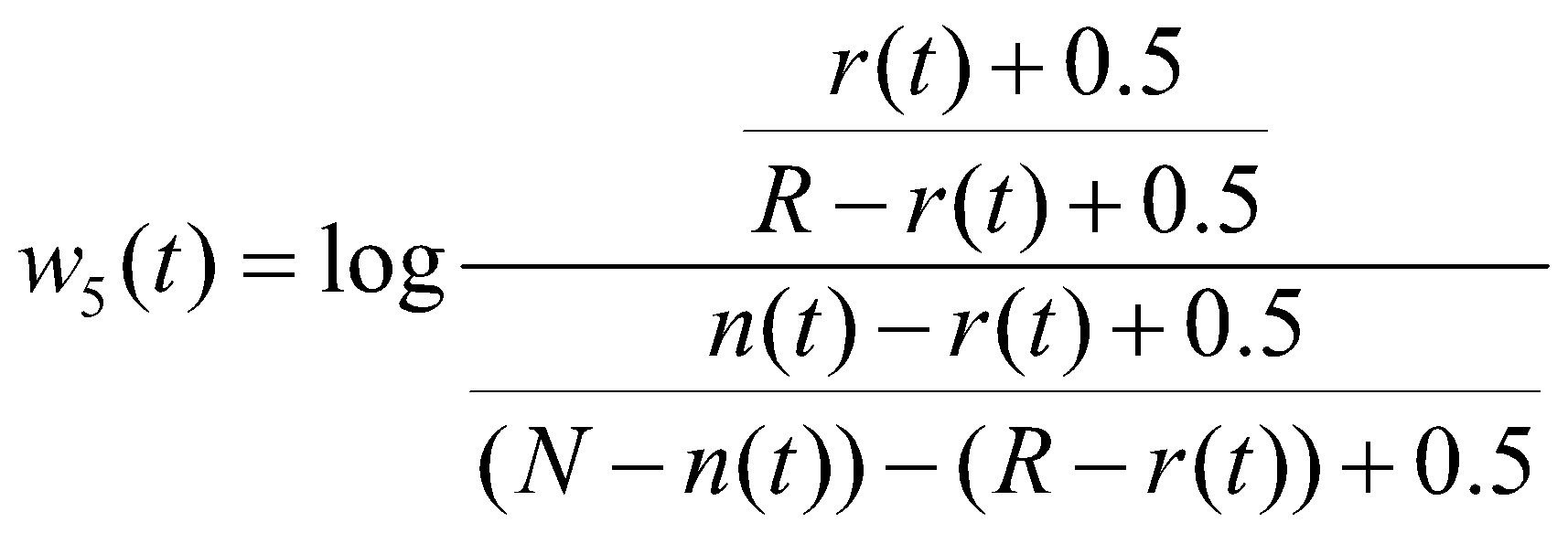
gamma = 1

Moreover, I use the TF-IDF scores from algorithm 1 in Q1.docx. To check whether a document is relevant or not, if the document D is relevant then {beta \* tf\_idf scores} is added to the alpha query vector, in else case {gamma \* tf\_idf scores} is added, when we find the final sum of all the vectors we get the updated query vector which is then used to rank the relevant scores and print to dat file.

**Algorithm 2**

**Probabilistic model**

In this model we use the following formula;



Where variables implicitly mean;

1. *N*=|***D***| be the total number of documents in the training set ***D***;
2. *R* be the number of relevant documents for a topic;
3. *n*(*t*) be the number of documents that contain term *t*; and
4. *r*(*t*) be the number of relevant documents that contain term *t*.

The resulting w5(t) is then used and we use the TF-IDF model to check whether a document is relevant or not, if the document D is relevant then {tf\_idf scores} is multiplied to the w5(t) vector, when we find the final sum of all the vectors we get the updated query vector which is then used to rank the relevant scores and print to dat file.